

MINIMUM SPANNING ACYCLE AND LIFETIME OF PERSISTENT HOMOLOGY IN THE LINIAL-MESHULAM PROCESS

YASUAKI HIRAOKA AND TOMOYUKI SHIRAI

ABSTRACT. This paper studies a higher dimensional generalization of Frieze's $\zeta(3)$ -limit theorem in the Erdős-Rényi graph process. Frieze's theorem states that the expected weight of the minimum spanning tree converges to $\zeta(3)$ as the number of vertices goes to infinity. In this paper, we study the d -Linial-Meshulam process as a model for random simplicial complexes, where $d = 1$ corresponds to the Erdős-Rényi graph process. First, we define spanning acycles as a higher dimensional analogue of spanning trees, and connect its minimum weight to persistent homology. Then, our main result shows that the expected weight of the minimum spanning acycle behaves in $O(n^{d-1})$.

Keywords. Random Simplicial Complex, Minimum Spanning Acycle, Linial-Meshulam Process, Persistent Homology

1. INTRODUCTION

Let $K_n = V_n \sqcup E_n$ be the complete graph with n vertices, where V_n and E_n are the sets of vertices and edges, respectively. We assign a uniform random variable $t_e \in [0, 1]$ independently for each edge $e \in E_n$, and define an increasing stochastic process of subgraphs of K_n by

$$(1.1) \quad K_n(t) = V_n \sqcup \{e \in E_n \mid t_e \leq t\}, \quad t \in [0, 1].$$

This process starts from V_n at time $t = 0$ and ends up with K_n at time $t = 1$. It is called the Erdős-Rényi graph process. By definition, $K_n(t)$ is equal in law to the Erdős-Rényi graph $G(n, t)$, which is obtained from K_n by retaining each edge with probability t and deleting it with probability $1 - t$ independently [5]. We also note that $K_n(t)$ defines a *random* filtration of K_n parametrized by $t \in [0, 1]$.

Let $\mathcal{S}^{(1)}$ be the set of spanning trees in K_n , i.e., the trees in K_n containing all vertices. Note that every spanning tree consists of $n - 1$ edges. The minimum spanning tree on K_n is defined as the spanning tree $T \in \mathcal{S}^{(1)}$ with the minimum weight $\text{wt}(T) = \sum_{e \in T} t_e$. Here, it is worth mentioning Kruskal's algorithm [15] for finding the minimum spanning tree. In Kruskal's algorithm, the weights $\{t_e\}_{e \in E_n}$ are treated as the birth times of edges. We start from the isolated vertices V_n at time 0, and then we expose an edge e at time t_e in order. If the edge e does not create a cycle, we keep it remained in our graph; otherwise we omit it. We repeat this procedure until the number of accepted edges becomes $n - 1$, and the derived tree will be the minimum spanning tree.

Frieze [7] shows the following significant result about the weight of the minimum spanning tree.

2010 *Mathematics Subject Classification.* 60C05, 05C80, 05E45.

Frieze's $\zeta(3)$ -Limit Theorem.

$$(1.2) \quad \mathbb{E}[\min_{T \in \mathcal{S}^{(1)}} \text{wt}(T)] \rightarrow \zeta(3) = 1.202 \dots$$

as $n \rightarrow \infty$, where $\zeta(s)$ is Riemann's zeta function.

This limit theorem has been investigated further in several directions, e.g., a central limit theorem and a tail estimate for the minimum weight, extensions to more general weight distributions and underlying graphs, and asymptotic expansions. Recent developments can be found in [2] and references therein. In the present paper, we will explore a higher dimensional generalization of this limit theorem.

One of the main ingredients in the proof of Frieze's theorem is the following formula connecting the weight of the minimum spanning tree to the integrated Betti number:

$$(1.3) \quad \min_{T \in \mathcal{S}^{(1)}} \text{wt}(T) = \int_0^1 \beta_0(t) dt.$$

Here, $\beta_0(t)$ is the reduced Betti number of $K_n(t)$, i.e., the rank of the reduced homology $H_0(K_n(t))$, and is equal to the number of connected components in $K_n(t)$ minus 1.

This formula is *deterministic* in the sense that it is valid for any realization of $\{t_e\}_{e \in E_n}$ and hence for the induced filtration $\{K_n(t)\}_{t \in [0,1]}$ of K_n . Given birth times $\{t_e\}_{e \in E_n}$, the reduced Betti number $\beta_0(t)$ decreases by 1 at time t_e if two connected components in $K_n(t_e -)$ are joined by adding the edge e . Since connected components can be regarded as generators of the 0-th homology, such a time t_e is viewed as a *death time* of the corresponding homology generator and the right-hand side of (1.3) gives the *lifetime sum* of $H_0(K_n(t))$. This observation naturally leads us to the notion of *persistent homology*.

The persistent homology [4, 23] (see Section 2.2 for details) has recently been studied as a tool to describe how topological features behave in a filtered topological space. In particular, it provides the concepts of the *birth* and *death times* of each topological feature, which measure the appearance and disappearance of the feature in the filtration. The *lifetime* is also defined as the difference between the birth and death times, and it measures the persistence of the feature in the filtration.

In this paper, we first show the following theorem as a higher dimensional extension of the formula (1.3), which is not just a counterpart of (1.3) but also sheds new light on the link among the lifetime sum of the persistent homology, the weights of the minimum spanning acycles, and the integrated Betti numbers.

Theorem 1.1. *Let X be a finite simplicial complex satisfying*

$$\beta_{d-1}(X^{(d)}) = \beta_{d-2}(X^{(d-1)}) = 0$$

for some $1 \leq d \leq \dim X$. Let $\mathcal{X} = \{X(t)\}_{t \in \mathbb{R}_{\geq 0}}$ be a filtration of X . Then, the following identities hold:

$$(1.4) \quad L_{d-1} = \min_{T \in \mathcal{S}^{(d)}} \text{wt}(T) - \max_{S \in \mathcal{S}^{(d-1)}} \text{wt}(X_{d-1} \setminus S)$$

$$(1.5) \quad = \int_0^\infty \beta_{d-1}(t) dt.$$

Here, $\mathbb{R}_{\geq 0}$ is the set of nonnegative reals, $\beta_k(t)$ is the k -th Betti number of $X(t)$, X_k is the set of k -simplices in X , $X^{(k)}$ is the k -dimensional skeleton of X , L_k is the lifetime sum of the k -th persistent homology (defined in (2.7)), and $\mathcal{S}^{(k)}$ is the set of k -spanning acycles (Definition 3.1). The proof of this theorem is given in Section 4.

The formula (1.3) is given as a special case $d = 1$ of this theorem. It should be remarked that, although only the death times are treated in the Erdős-Rényi graph process ($d = 1$), we also need to study the birth times in the higher dimensional case. This effect causes the second term in (1.4), and the formulation using the birth and death times in persistent homology fits this extension well. Furthermore, we remark that Frieze's theorem can also be expressed by using the lifetime sum as follows:

$$\mathbb{E}[L_0] \rightarrow \zeta(3) \quad \text{as } n \rightarrow \infty.$$

Based on these formulae (1.4) and (1.5), we study a higher dimensional generalization of the Erdős-Rényi graph as random simplicial complexes. The connectivity and acyclicity of graphs, which are commonly studied in random graphs, can be interpreted by using 0-th and 1-st homologies, respectively. Then, it is natural to generalize classical results in the Erdős-Rényi graph into analogues expressed by higher dimensional homology of suitable random simplicial complexes (e.g., see the papers [12, 13, 16, 18, 21] and references therein for recent topics of random simplicial complexes).

In this paper, we consider two processes of random simplicial complexes, the Linial-Meshulam process [16] and clique complex process [12], both of which can be regarded as natural generalizations of the Erdős-Rényi graph process. Precise definitions of these processes are given in Section 5. Our main result shows the following higher dimensional generalization of Frieze's $\zeta(3)$ -limit theorem in the Linial-Meshulam process.

Theorem 1.2. *Let L_{d-1} be the lifetime sum of the $(d-1)$ -st persistent homology of the d -Linial-Meshulam process ($d \geq 1$) on n -vertices. Then,*

$$(1.6) \quad \mathbb{E}[L_{d-1}] = O(n^{d-1})$$

as $n \rightarrow \infty$.

For $d = 1$, we already know that the limiting value is $\zeta(3)$ from Frieze's theorem and this agrees with (1.6).

This paper is organized as follows. The fundamental concepts of homology and persistent homology are explained in Section 2. Here, the algebraic formulation using graded modules and the analytic formulation using counting measures are introduced for persistent homology, and both formulations are used to derive Theorem 1.1 and Theorem 1.2. In Section 3, we summarize a determinantal formula of boundary maps by means of spanning acycles. Section 4 is devoted to proving Theorem 1.1. In Section 5, we explain random persistence diagrams as point processes, and then introduce the d -Linial-Meshulam process and the clique complex process. The proof of Theorem 1.2 is presented in Section 6. Furthermore, we also show a partial result (Theorem 6.10) on the higher dimensional extension of Frieze's theorem for the clique complex process. In Section 7, we list some conjectures and open questions.

2. PERSISTENT HOMOLOGY AND LIFETIME

2.1. Homology. We first recall some fundamental concepts of simplicial homology. For more details, the reader may refer to [8]. Let X be a simplicial complex on a finite set $V = \{1, \dots, n\}$, i.e., a collection of nonempty subsets of V which includes all elements in V and is closed under the operation of taking nonempty subsets. An element $\sigma \in X$ with $|\sigma| = k + 1$ is called a k -simplex and k is called its dimension. The dimension $\dim X$ of the simplicial complex X is given by the maximum dimension of simplices in X . We denote the set of k -simplices in X and its cardinality by X_k and $f_k(X) = |X_k|$, respectively. The k -dimensional skeleton of X is defined by $X^{(k)} = \bigsqcup_{j=0}^k X_j$. In this paper, we only deal with finite simplicial complexes, i.e., $|V| < \infty$.

For a simplicial complex X , the boundary map $\partial_k : C_k(X) \rightarrow C_{k-1}(X)$ and the chain complex

$$(2.1) \quad \cdots \longrightarrow C_{k+1}(X) \xrightarrow{\partial_{k+1}} C_k(X) \xrightarrow{\partial_k} C_{k-1}(X) \longrightarrow \cdots$$

in the integer coefficient are defined in a standard way. For $\sigma = \{v_0, \dots, v_k\} \in X$, we set its oriented simplex by the ordering $v_0 < \dots < v_k$ and denote it by $\langle \sigma \rangle = \langle v_0 \cdots v_k \rangle$. Then, the k -th homology $H_k(X) = Z_k(X)/B_k(X)$ is defined as the quotient \mathbb{Z} -module of $Z_k(X) = \ker \partial_k$ and $B_k(X) = \text{im } \partial_{k+1}$.

In this paper, we use the reduced homology $\tilde{H}_0(X)$ for $k = 0$, which is given by $H_0(X) \simeq \tilde{H}_0(X) \oplus \mathbb{Z}$. For simplicity, we use the same symbol $H_0(X)$ for the 0-th reduced homology and omit to specify “reduced” from now on. We also note that the homology can be represented as $H_k(X) \simeq T_k(X) \oplus \mathbb{Z}^{\beta_k(X)}$, where $T_k(X)$ and $\beta_k(X)$ are called the k -th torsion and the k -th Betti number, respectively.

For simplicial complexes $Y \subset X$, let $C_k(X, Y) = C_k(X)/C_k(Y)$ be the quotient module. The boundary map in the chain complex (2.1) naturally induces the relative chain complex

$$(2.2) \quad \cdots \longrightarrow C_{k+1}(X, Y) \xrightarrow{\partial_{k+1}} C_k(X, Y) \xrightarrow{\partial_k} C_{k-1}(X, Y) \longrightarrow \cdots$$

Then, the k -th relative homology $H_k(X, Y)$ is defined by the same way as $H_k(X, Y) = Z_k(X, Y)/B_k(X, Y)$, where $Z_k(X, Y) = \ker \partial_k$ and $B_k(X, Y) = \text{im } \partial_{k+1}$ in (2.2). It is well known that there exists an exact sequence for a pair $Y \subset X$:

$$(2.3) \quad \cdots \longrightarrow H_{k+1}(X, Y) \longrightarrow H_k(Y) \longrightarrow H_k(X) \longrightarrow H_k(X, Y) \longrightarrow H_{k-1}(Y) \longrightarrow \cdots$$

2.2. Persistent Homology. Let $\mathbb{Z}_{\geq 0}$ and $\mathbb{R}_{\geq 0}$ be the sets of nonnegative integers and reals, respectively. Let $\mathcal{X} = \{X(t) \mid t \in \mathbb{R}_{\geq 0}\}$ be a right continuous filtration of a simplicial complex X . Namely, $X(t)$ is a subcomplex of X , $X(t) \subset X(t')$ for $t \leq t'$, and $X(t) = \bigcap_{t < t'} X(t')$. We assume that there exists a saturation time T such that $X(T) = X$. For each simplex $\sigma \in X$, let $t_\sigma = \min\{t \in \mathbb{R}_{\geq 0} \mid \sigma \in X(t)\}$ denote the birth time of σ .

Let K be a field of characteristic zero, and let $K[\mathbb{R}_{\geq 0}]$ be a monoid ring. That is, $K[\mathbb{R}_{\geq 0}]$ is a K -vector space of formal linear combinations of elements in $\mathbb{R}_{\geq 0}$ equipped with a ring structure

$$(at) \cdot (bs) = (ab)(t + s), \quad a, b \in K, \quad t, s \in \mathbb{R}_{\geq 0}.$$

In the following, the elements in $K[\mathbb{R}_{\geq 0}]$ are expressed by linear combinations of (formal) monomials az^t , where $a \in K$, $t \in \mathbb{R}_{\geq 0}$, and z is an indeterminate. Then, the product of two elements are given by the linear extension of $az^t \cdot bz^s = abz^{t+s}$.

For $t \in \mathbb{R}_{\geq 0}$, let $C_k(X(t))$ be the K -vector space spanned by the oriented k -simplices in $X(t)$. The k -th chain group $C_k(\mathcal{X})$ of \mathcal{X} is defined as a graded module over the monoid ring $K[\mathbb{R}_{\geq 0}]$ by taking a direct sum

$$C_k(\mathcal{X}) = \bigoplus_{t \in \mathbb{R}_{\geq 0}} C_k(X(t)) = \{(c_t) \mid c_t \in C_k(X(t)), t \in \mathbb{R}_{\geq 0}\},$$

where the action of a monomial z^s on $C_k(\mathcal{X})$ is given by the right shift operator

$$z^s \cdot (c_t) = (c'_t), \quad c'_t = \begin{cases} c_{t-s}, & t \geq s \\ 0, & t < s \end{cases}.$$

For an oriented simplex $\langle \sigma \rangle$, let us define

$$\langle\langle \sigma \rangle\rangle = (c_t), \quad c_t = \begin{cases} \langle \sigma \rangle, & t = t_\sigma \\ 0, & t \neq t_\sigma \end{cases}.$$

Then, the set $\Xi_k = \{\langle\langle \sigma \rangle\rangle \mid \sigma \in X_k\}$ forms a basis of $C_k(\mathcal{X})$. The boundary map $\delta_k : C_k(\mathcal{X}) \rightarrow C_{k-1}(\mathcal{X})$ is defined by the linear extension of

$$(2.4) \quad \delta_k \langle\langle \sigma \rangle\rangle = \sum_{j=0}^k (-1)^j z^{t_\sigma - t_{\sigma_j}} \langle\langle \sigma_j \rangle\rangle,$$

where $\langle \sigma \rangle = \langle v_0 \cdots v_k \rangle$ and $\sigma_j = \sigma \setminus \{v_j\}$. We note $t_\sigma - t_{\sigma_j} \geq 0$ from $\sigma_j \subset \sigma$. The matrix form of δ_k using the standard bases Ξ_k and Ξ_{k-1} consists of entries $\pm z^t \in K[\mathbb{R}_{\geq 0}]$.

The cycle group $Z_k(\mathcal{X})$ and the boundary group $B_k(\mathcal{X})$ in $C_k(\mathcal{X})$ are defined by

$$Z_k(\mathcal{X}) = \ker \delta_k, \quad B_k(\mathcal{X}) = \operatorname{im} \delta_{k+1}.$$

It follows from $\delta_k \circ \delta_{k+1} = 0$ that $B_k(\mathcal{X}) \subset Z_k(\mathcal{X})$. Then, the k -th persistent homology is defined by

$$H_k(\mathcal{X}) = Z_k(\mathcal{X}) / B_k(\mathcal{X}).$$

We note that the persistent homology is a graded module over $K[\mathbb{R}_{\geq 0}]$.

The following theorem is known as the structure theorem of the persistent homology.

Theorem 2.1 ([23]). *There uniquely exist indices $p, q \in \mathbb{Z}_{\geq 0}$ and $(b_i, d_i) \in \mathbb{R}_{\geq 0}^2$ for $i = 1, \dots, p$ with $b_i < d_i$ and $b_i \in \mathbb{R}_{\geq 0}$ for $i = p+1, \dots, p+q$ such that the following isomorphism holds:*

$$(2.5) \quad H_k(\mathcal{X}) \simeq \bigoplus_{i=1}^p \left((z^{b_i}) / (z^{d_i}) \right) \oplus \bigoplus_{i=p+1}^{p+q} (z^{b_i}),$$

where (z^a) expresses an ideal in $K[\mathbb{R}_{\geq 0}]$ generated by the monomial z^a . When p or q is zero, the corresponding direct sum is ignored.

Here b_i and d_i are called the birth and the death times, respectively, and they measure the events of appearance and disappearance of topological features in the filtration \mathcal{X} . Namely, it expresses that a homology generator is born in $H_k(X(b_i))$, persists in $H_k(X(t))$ for $b_i \leq t \leq d_i$, and dies in $H_k(X(d_i))$. The lifetime l_i of the pair (b_i, d_i) is defined by $l_i = d_i - b_i$. For $p+1 \leq i \leq p+q$, we assign the death time

$d_i = \infty$ as the element of the extended nonnegative reals $\overline{\mathbb{R}}_{\geq 0} := \mathbb{R}_{\geq 0} \cup \{\infty\}$. We remark that the representation (2.5) of the persistent homology is the counterpart to the one using the torsion and free modules in the standard homology. Both are derived from the structure theorem of finitely generated modules over PID.

The indecomposable decomposition (2.5) of the persistent homology can be expressed by using a multiset called the k -th persistence diagram

$$(2.6) \quad D_k(\mathcal{X}) = \{(b_i, d_i) \in \overline{\mathbb{R}}_{\geq 0}^2 \mid i = 1, \dots, p+q\}.$$

Similar to homology, we use the reduced persistent homology for $k = 0$, which is defined by deleting one generator with infinite death time from $H_0(\mathcal{X})$. For simplicity, we use the same symbol $H_0(\mathcal{X})$ and omit to specify “reduced”. The persistence diagram $D_0(\mathcal{X})$ is also defined in a reduced sense.

2.3. Lifetime Formula I. We denote the lifetime sum of the k -th persistent homology by

$$(2.7) \quad L_k = \sum_{i=1}^{p+q} (d_i - b_i),$$

where L_k is understood as ∞ when $q \geq 1$.

It is often convenient to regard the k -th persistence diagram (2.6) as a counting measure

$$\xi_k = \sum_{0 \leq x < y \leq \infty} m_{(x,y)} \delta_{(x,y)}$$

on the set $\Delta = \{(x, y) \in \overline{\mathbb{R}}_{\geq 0}^2 \mid x \leq y\}$, where $\delta_{(x,y)}$ is the delta measure at (x, y) and

$$m_{(x,y)} = |\{1 \leq i \leq p+q \mid (b_i, d_i) = (x, y)\}|$$

is the multiplicity. We note that

$$(2.8) \quad \beta_k(t) = \xi_k([0, t] \times [t, \infty)),$$

where $\beta_k(t) = \beta_k(X(t))$.

We write

$$\langle \xi_k, f \rangle = \int_{\Delta} f(x, y) \xi_k(dxdy)$$

for any measurable function $f : \Delta \rightarrow \mathbb{R}$ as long as the right-hand side makes sense. For example, when f is the indicator function I_A of a measurable set $A \subset \Delta$, $\langle \xi_k, I_A \rangle = \xi_k(A)$ is the number of points inside A counted with multiplicity. By setting $f(x, y) = y - x$, we also have

$$(2.9) \quad \langle \xi_k, f \rangle = \int_{\Delta} (y - x) \xi_k(dxdy) = \sum_{i=1}^{p+q} (d_i - b_i) = L_k.$$

Then, we easily obtain the following formula of the lifetime sum.

Proposition 2.2.

$$L_k = \int_{[0, \infty]} \beta_k(t) dt.$$

Proof. By Fubini's theorem, (2.8) and (2.9), we see that

$$\begin{aligned}
 L_k &= \int_{\Delta} (y-x) \xi_k(dxdy) \\
 &= \int_{\Delta} \xi_k(dxdy) \int_{[0,\infty]} I(0 \leq x \leq t \leq y \leq \infty) dt \\
 &= \int_{[0,\infty]} dt \int_{\Delta} I_{[0,t]}(x) I_{[t,\infty]}(y) \xi_k(dxdy) \\
 &= \int_{[0,\infty]} \beta_k(t) dt.
 \end{aligned}$$

When $q \geq 1$, the both sides are ∞ . \square

The persistent homology treated in this paper does not have the latter part in the indecomposable decomposition (2.5). Hence, we always suppose the case $q = 0$ from now on.

Remark 2.3. This lifetime formula can be regarded as Little's formula in queuing theory [19].

Remark 2.4. The lifetime sum L_k can be regarded as the ℓ^1 -norm $\|\vec{l}\|_1$ of a sequence $\vec{l} = (l_i)_{i=1}^p$ of lifetimes $l_i = d_i - b_i$ in the k -th persistent homology. P. Bubenik points out that the squared ℓ^2 -norm of \vec{l} is equal to 4 times the L^1 -norm of the persistent landscape [1]. In order to make clear the connection to the persistent landscape, we derive a similar integral formula for the ℓ^2 -norm.

First, let us define the $(t-s)$ -persistent homology [4]

$$H_k(s, t) = Z_k(X(s)) / (B_k(X(t)) \cap Z_k(X(s))).$$

We note that

$$\text{rank } H_k(s, t) = \int_{\Delta} I_{[0,s]}(x) I_{[t,\infty]}(y) \xi_k(dxdy),$$

and we denote the left-hand side by $\beta_k(s, t)$. Then, the integral formula for the ℓ^2 -norm is given by

$$\|\vec{l}\|_2^2 = 2 \int_{0 \leq s \leq t \leq \infty} \beta_k(s, t) ds dt.$$

This formula is derived in a similar way using Fubini's theorem:

$$\begin{aligned}
 \|\vec{l}\|_2^2 &= \int_{\Delta} (y-x)^2 \xi_k(dxdy) \\
 &= \int_{\Delta} \xi_k(dxdy) \left(\int_{[0,\infty]} I(0 \leq x \leq t \leq y \leq \infty) dt \right)^2 \\
 &= \int_{[0,\infty]^2} dt ds \int_{\Delta} I_{[0,t]}(x) I_{[t,\infty]}(y) I_{[0,s]}(x) I_{[s,\infty]}(y) \xi_k(dxdy) \\
 &= \int_{[0,\infty]^2} dt ds \int_{\Delta} I_{[0,t \wedge s]}(x) I_{[t \vee s, \infty]}(y) \xi_k(dxdy) \\
 &= 2 \int_{0 \leq s \leq t \leq \infty} \int_{\Delta} I_{[0,s]}(x) I_{[t,\infty]}(y) \xi_k(dxdy) \\
 &= 2 \int_{0 \leq s \leq t \leq \infty} \beta_k(s, t) ds dt,
 \end{aligned}$$

where $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

3. SPANNING ACYCLE AND DETERMINANTAL FORMULA

In this section, we basically follow the argument in [3].

3.1. Spanning Acycle. Let X be a simplicial complex and let $k \in \mathbb{N}$ be $k \leq \dim X$. For a subset $S \subset X_k$, we define a k -dimensional subcomplex of X by

$$(3.1) \quad X_S = S \sqcup X^{(k-1)}.$$

Definition 3.1. A subset $S \subset X_k$ is called a k -spanning acycle if

- (a) $H_k(X_S) = 0$, and
- (b) $|H_{k-1}(X_S)| < \infty$.

The set of k -spanning acycles in X is denoted by $\mathcal{S}^{(k)}$.

This definition is a natural generalization of the spanning trees of a graph. For $\dim X = 1$ and $k = 1$, S is a subset of edges and $X_S = V \sqcup S$ is a graph. In this case, the conditions (a) and (b) are equivalent that X_S has no cycles and X_S is connected, respectively. This means that the 1-spanning acycle S is nothing but a spanning tree.

Remark 3.2. This definition is originally introduced by Kalai [14] for X being a k -dimensional simplicial complex with the complete $(k-1)$ -skeleton. This is essentially the same as *simplicial spanning tree* given in [3].

Example 3.3. Let σ be a 3-simplex and let X be the simplicial complex consisting of all proper subsets in σ . Then, any collections of three 2-simplices in X become 2-spanning acycles. On the other hand, any collections of two 2-simplices are not 2-spanning acycles. More generally, the set of the 2-simplices in a 2-dimensional triangulated sphere minus one 2-simplex forms a 2-spanning acycle.

Lemma 3.4. *If there exists a k -spanning acycle S in X , then $|H_{k-1}(X^{(k)})| < \infty$.*

Proof. It follows from (3.1) that $C_k(X_S) \subset C_k(X)$ and $C_j(X_S) = C_j(X)$ for $j < k$. Hence, we have $\text{im } \partial_k|_S \subset \text{im } \partial_k$ and $\ker \partial_{k-1}|_S = \ker \partial_{k-1}$, where $\partial_k|_S$ expresses the restriction of ∂_k on $C_k(X_S)$. This implies that there exists a surjection from $H_{k-1}(X_S)$ to $H_{k-1}(X^{(k)})$. Hence, if S is a k -spanning acycle, the condition (b) implies $|H_{k-1}(X^{(k)})| < \infty$. \square

Remark 3.5. In [20], instead of k -spanning acycles, the notion of k -bases is considered in the context of matroids. In that definition, for example, the set of the 2-simplices in a 2-dimensional triangulated oriented surface with genus $g \geq 1$ minus one 2-simplex forms a 2-base, whereas there are no 2-spanning acycles in our definition from Lemma 3.4.

For $k \geq 0$, let us define

$$\gamma_k(X) = f_k(X^{(k)}) - \beta_k(X^{(k)}) + \beta_{k-1}(X^{(k)})$$

with $\beta_{-1}(X^{(0)}) = 0$. Then, we obtain a complementary characterization of k -spanning acycles as follows.

Lemma 3.6. *Any two of the three conditions (a), (b) in Definition 3.1 and*

$$(3.2) \quad |S| = \gamma_k(X)$$

imply the third.

Proof. First, we note from (3.1) that

$$\begin{aligned} f_j(X_S) &= f_j(X^{(k)}), & 0 \leq j \leq k-1, \\ \beta_j(X_S) &= \beta_j(X^{(k)}), & 0 \leq j \leq k-2. \end{aligned}$$

By the Euler-Poincaré formula, we see that

$$\begin{aligned} \chi(X_S) - \chi(X^{(k)}) &= (-1)^k \{f_k(X_S) - f_k(X^{(k)})\} \\ &= \sum_{j=k-1}^k (-1)^j \{\beta_j(X_S) - \beta_j(X^{(k)})\}, \end{aligned}$$

where χ is the Euler characteristic. This is equivalent to

$$\{|S| - \gamma_k(X)\} + \beta_{k-1}(X_S) - \beta_k(X_S) = 0,$$

and the assertion is obvious from this identity. \square

The cardinality of a k -spanning acycle is given as follows.

Corollary 3.7. *If S is a k -spanning acycle, then*

$$(3.3) \quad |S| = \gamma_k(X) = f_k(X^{(k)}) - \beta_k(X^{(k)}).$$

Proof. The claim follows from Lemma 3.4 and Lemma 3.6. \square

Lemma 3.8. *For $k \geq 0$,*

$$\gamma_k(X) = \dim \ker \partial_{k-1} - \delta_{k,1} + \delta_{k,0},$$

and for $k \geq 1$,

$$(3.4) \quad f_{k-1}(X) - \gamma_k(X) = \gamma_{k-1}(X) - \beta_{k-2}(X^{(k-1)}).$$

Proof. Set $N_k = \dim \ker \partial_k$ and $I_k = \dim \operatorname{im} \partial_k$. Then, we have

$$\begin{aligned} \gamma_k(X) &= (N_k + I_k) - (N_k - \delta_{k,0}) + (N_{k-1} - \delta_{k,1} - I_k) \\ &= N_{k-1} - \delta_{k,1} + \delta_{k,0}. \end{aligned}$$

Similarly, for $k \geq 1$,

$$\begin{aligned} &f_{k-1}(X) - \gamma_k(X) - \gamma_{k-1}(X) + \beta_{k-2}(X^{(k-1)}) \\ &= (N_{k-1} + I_{k-1}) - (N_{k-1} - \delta_{k,1} + \delta_{k,0}) \\ &\quad - (N_{k-2} - \delta_{k-1,1} + \delta_{k-1,0}) + (N_{k-2} - I_{k-1} - \delta_{k-2,0}) \\ &= -\delta_{k,0} = 0. \end{aligned}$$

\square

Proposition 3.9. *Let X be a simplicial complex satisfying*

$$(3.5) \quad \beta_{j-1}(X^{(j)}) = 0, \quad 1 \leq j < \dim X.$$

Then, for $0 \leq k \leq \dim X$,

$$(3.6) \quad \gamma_k(X) = (-1)^k \left\{ 1 - \sum_{j=0}^{k-1} (-1)^j f_j(X) \right\}.$$

Proof. The equality obviously holds for $k = 0$. From (3.4) and (3.5), we have

$$f_{j-1}(X) = \gamma_j(X) + \gamma_{j-1}(X)$$

for $1 \leq j \leq \dim X$. Taking the alternating sum of the above leads to

$$\sum_{j=1}^k (-1)^{j-1} f_{j-1}(X) = \gamma_0(X) - (-1)^k \gamma_k(X),$$

and (3.6) follows from this. \square

Example 3.10. Let X be the $(n-1)$ -dimensional maximal simplicial complex on n vertices. It is obvious that $H_{k-1}(X^{(k)}) = 0$ (and thus $\beta_{k-1}(X^{(k)}) = 0$) for $k = 1, 2, \dots, n-1$ and $f_k(X) = \binom{n}{k+1}$. Then, it follows from (3.6) that

$$\gamma_k(X) = \binom{n-1}{k}$$

for $k = 0, 1, \dots, n-1$.

3.2. Determinantal Formula. Let $d \in \mathbb{N}$ be $d \leq \dim X$. Let us express the boundary map $\partial_d : C_d(X) \rightarrow C_{d-1}(X)$ in the matrix form under the standard bases (the sets of oriented simplices). For $K \subset X_{d-1}$ and $S \subset X_d$, we denote by ∂_{KS} the submatrix of ∂_d restricted to the rows and columns spanned by the simplices in K and S , respectively. The submatrices ∂_K and ∂_S are similarly defined.

Lemma 3.11. *Let $K, L \subset X_{d-1}$ with $K = X_{d-1} \setminus L$ and $S \subset X_d$. Suppose that $|K| = |S| = \gamma_d(X)$. Then, $\det \partial_{KS} \neq 0$ if and only if $S \in \mathcal{S}^{(d)}$ and $H_{d-1}(X_L) = 0$. In this case,*

$$|\det \partial_{KS}| = |H_{d-1}(X_S, X_L)|.$$

Proof. It follows from (3.1) that X_L is a subcomplex of X_S and

$$(X_S)_k = (X_L)_k = X_k, \quad 0 \leq k \leq d-2.$$

This implies $H_k(X_S, X_L) = 0$ for $0 \leq k \leq d-2$. Then, we have an exact sequence (see (2.3))

$$(3.7) \quad 0 \rightarrow H_d(X_S) \rightarrow H_d(X_S, X_L) \rightarrow H_{d-1}(X_L) \rightarrow H_{d-1}(X_S) \rightarrow \cdots.$$

Suppose that $H_d(X_S) = H_{d-1}(X_L) = 0$. Then, the exact sequence (3.7) leads to

$$\ker \partial_{KS} = H_d(X_S, X_L) = 0,$$

which implies $\det \partial_{KS} \neq 0$.

Assume to the contrary that $\det \partial_{KS} \neq 0$. Then, $H_d(X_S, X_L) = \ker \partial_{KS} = 0$, and hence $H_d(X_S) = 0$ from (3.7). This together with $|S| = \gamma_d(X)$ means $S \in \mathcal{S}^{(d)}$ from Lemma 3.6, and $H_{d-1}(X_S)$ is a finite group. Furthermore, (3.7) leads to an injection

$$0 \rightarrow H_{d-1}(X_L) \rightarrow H_{d-1}(X_S).$$

Because of $\dim X_L = d-1$, $H_{d-1}(X_L)$ is free. Thus, $H_{d-1}(X_L)$ must be zero.

It follows from $C_{d-2}(X_S, X_L) = 0$ that

$$H_{d-1}(X_S, X_L) = C_{d-1}(X_S, X_L) / \text{im } \partial_{KS}.$$

Hence, $\ker \partial_{KS} = 0$ implies that $H_{d-1}(X_S, X_L)$ is a finite group of order $|\det \partial_{KS}|$. \square

Remark 3.12. Let K and L be as in Lemma 3.11. It follows from (3.4) that

$$\begin{aligned} |L| &= f_{d-1}(X) - \gamma_d(X) \\ &= \gamma_{d-1}(X) - \beta_{d-2}(X^{(d-1)}). \end{aligned}$$

Hence, we need the condition $\beta_{d-2}(X^{(d-1)}) = 0$ for $L \in \mathcal{S}^{(d-1)}$ as stated in Lemma 3.4.

Corollary 3.13. Let X be a simplicial complex with $\beta_{d-2}(X^{(d-1)}) = 0$. Suppose $K \subset X_{d-1}$ and $S \subset X_d$ satisfy $|K| = |S| = \gamma_d(X)$. Then, $\det \partial_{KS} \neq 0$ if and only if $S \in \mathcal{S}^{(d)}$ and $X_{d-1} \setminus K \in \mathcal{S}^{(d-1)}$.

Proof. The assertion immediately follows from Lemma 3.11 and Remark 3.12. \square

Let $\mathbf{x} = (x_\sigma)_{\sigma \in X_{d-1}}$ and $\mathbf{y} = (y_\eta)_{\eta \in X_d}$ be indeterminates corresponding to the $(d-1)$ -simplices and the d -simplices in X , respectively. Set

$$(3.8) \quad \partial_d(\mathbf{x}, \mathbf{y}) = \text{diag}(\mathbf{x}) \partial_d \text{diag}(\mathbf{y}),$$

where $\text{diag}(\mathbf{x})$ is the diagonal matrix with entries being \mathbf{x} .

Proposition 3.14. Let $K \subset X_{d-1}$ with $|K| = \gamma_d(X)$. Then,

$$\det \partial_d(\mathbf{x}, \mathbf{y})_K \partial_d(\mathbf{x}, \mathbf{y})_K^t = \sum_{S \in \mathcal{S}^{(d)}} (\det \partial_{KS})^2 \mathbf{x}_K^2 \mathbf{y}_S^2,$$

where $\mathbf{x}_K = \prod_{\sigma \in K} x_\sigma$ and $\mathbf{y}_S = \prod_{\eta \in S} y_\eta$

Proof. The Binet-Cauchy formula leads to

$$\begin{aligned} \det \partial_d(\mathbf{x}, \mathbf{y})_K \partial_d(\mathbf{x}, \mathbf{y})_K^t &= \sum_{\substack{S \subset X_d \\ |S| = \gamma_d(X)}} (\det \partial_{KS})^2 \mathbf{x}_K^2 \mathbf{y}_S^2 \\ &= \sum_{S \in \mathcal{S}^{(d)}} (\det \partial_{KS})^2 \mathbf{x}_K^2 \mathbf{y}_S^2. \end{aligned}$$

The second equality follows from Lemma 3.11. \square

Lemma 3.15. Suppose that $S \in \mathcal{S}^{(d)}$ and $L \in \mathcal{S}^{(d-1)}$, and set $K = X_{d-1} \setminus L$. Then,

$$(3.9) \quad |\det \partial_{KS}| = \frac{|H_{d-1}(X_S)| \cdot |H_{d-2}(X_L)|}{|H_{d-2}(X_S)|}.$$

Proof. It follows from $L \in \mathcal{S}^{(d-1)}$ that we have an exact sequence

$$0 \rightarrow H_{d-1}(X_S) \rightarrow H_{d-1}(X_S, X_L) \rightarrow H_{d-2}(X_L) \rightarrow H_{d-2}(X_S) \rightarrow 0.$$

Then, $H_{d-2}(X_S)$ and $H_{d-1}(X_S, X_L)$ are finite. Therefore, we have

$$|\det \partial_{KS}| = |H_{d-1}(X_S, X_L)| = \frac{|H_{d-1}(X_S)| \cdot |H_{d-2}(X_L)|}{|H_{d-2}(X_S)|}.$$

\square

From this lemma and Proposition 3.14, we have the following theorem.

Theorem 3.16. *Suppose $K \subset X_{d-1}$ with $|K| = \gamma_d(X)$ and $L = X_{d-1} \setminus K \in \mathcal{S}^{(d-1)}$. Then,*

$$(3.10) \quad \det \partial_d(\mathbf{x}, \mathbf{y})_K \partial_d(\mathbf{x}, \mathbf{y})_K^t = \sum_{S \in \mathcal{S}^{(d)}} \left(\frac{|H_{d-1}(X_S)| \cdot |H_{d-2}(X_L)|}{|H_{d-2}(X_S)|} \right)^2 \mathbf{x}_K^2 \mathbf{y}_S^2.$$

Example 3.17. Let X be a triangulation of a 2-dimensional sphere. For $X_1 \setminus K \in \mathcal{S}^{(1)}$, we have

$$\det \partial_2(\mathbf{x}, \mathbf{y})_K \partial_2(\mathbf{x}, \mathbf{y})_K^t = \mathbf{x}_K^2 \sum_{S \in \mathcal{S}^{(2)}} |H_1(X_S)|^2 \mathbf{y}_S^2 = \mathbf{x}_K^2 \sum_{S \in \mathcal{S}^{(2)}} \mathbf{y}_S^2.$$

Example 3.18. Let X be the $(n-1)$ -dimensional maximal simplicial complex on n vertices and L be a set of $(d-1)$ -simplices ($d < n$) in X with one fixed vertex. Let us set $\mathbf{x} = (x_\sigma)$ and $\mathbf{y} = (y_\eta)$ to be $x_\sigma = 1$ and $y_\eta = 1$ for all $\sigma \in X_{d-1}$ and $\eta \in X_d$. Then, Theorem 3.16 is reduced to the Kalai's result [14]. Namely, because of $H_{d-2}(X_L) = 0$ and $H_{d-2}(X_S) = 0$ in this setting, the equality (3.10) becomes

$$n \binom{n-2}{d} = \sum_{S \in \mathcal{S}^{(d)}} |H_{d-1}(X_S)|^2.$$

Here, the left-hand side is derived by showing that the eigenvalues of $\partial_d(\mathbf{x}, \mathbf{y})_K \partial_d(\mathbf{x}, \mathbf{y})_K^t$ are given by 1 and n with multiplicities $\binom{n-2}{d-1}$ and $\binom{n-2}{d}$, respectively. The case $d = 1$ is the Cayley's formula counting the number of spanning trees.

4. LIFETIME FORMULA II

In this section, we give a proof of Theorem 1.1. Throughout this section, let us set $d \in \mathbb{N}$ as $d \leq \dim X$. Furthermore, we assume that the simplicial complex X satisfies

$$\beta_{d-1}(X^{(d)}) = \beta_{d-2}(X^{(d-1)}) = 0.$$

Let $\mathcal{X} = \{X(t) \mid t \in \mathbb{R}_{\geq 0}\}$ be a filtration of X . A minimum d -spanning acycle of the filtration \mathcal{X} is defined as a spanning acycle $S \in \mathcal{S}^{(d)}$ with the minimum weight $\text{wt}(S) = \sum_{\sigma \in S} t_\sigma$ among $\mathcal{S}^{(d)}$, where t_σ is the birth time of the simplex σ .

We denote by M the matrix form of the d -th boundary map δ_d of the persistent homology $H_*(\mathcal{X})$ under the standard bases Ξ_d, Ξ_{d-1} . We also denote its evaluation at $z = 1$ by $D = M|_{z=1}$, which is a matrix form of ∂_d . It should be noted that

$$\text{rank } \delta_d = \text{rank } \partial_d = f_d(X) - \ker \partial_d = f_d(X^{(d)}) - \beta_d(X^{(d)}) = \gamma_d(X).$$

Let us denote the elementary divisors of M by $d_1 = z^{e_1}, \dots, d_r = z^{e_r}$, where $r = \gamma_d(X)$.

Proposition 4.1. *Let $K \subset X_{d-1}$ with $|K| = \gamma_d(X)$. Then,*

$$(4.1) \quad \det M_K M_K^t = z^{2e(K)} \sum_{S \in \mathcal{S}^{(d)}} (\det D_{KS})^2 z^{2\tau(S)},$$

where

$$\tau(S) = \text{wt}(S) - \min_{S' \in \mathcal{S}^{(d)}} \text{wt}(S'), \quad e(K) = \min_{S \in \mathcal{S}^{(d)}} \text{wt}(S) - \text{wt}(K),$$

and $e(K)$ is nonnegative.

Proof. By setting $\mathbf{x} = (z^{-t_\sigma})_{\sigma \in X_{d-1}}$ and $\mathbf{y} = (z^{t_\eta})_{\eta \in X_d}$, $\partial_d(\mathbf{x}, \mathbf{y})$ defined in (3.8) coincides with $\delta_d : C_d(\mathcal{X}) \rightarrow C_{d-1}(\mathcal{X})$. By Proposition 3.14, we obtain

$$\begin{aligned} \det M_K M_K^t &= \sum_{S \in \mathcal{S}^{(d)}} (\det M_{KS})^2 \\ &= \sum_{S \in \mathcal{S}^{(d)}} (\det D_{KS})^2 z^{2(\text{wt}(S) - \text{wt}(K))} \\ &= z^{2e(K)} \sum_{S \in \mathcal{S}^{(d)}} (\det D_{KS})^2 z^{2\tau(S)}. \end{aligned}$$

The claim $e(K) \geq 0$ follows from the fact $t_\sigma \leq t_\eta$ for $\sigma \subset \eta$. \square

Lemma 4.2. *For the elementary divisors $d_1 = z^{e_1}, \dots, d_r = z^{e_r}$ of M ,*

$$\min_{K \in \mathcal{S}_c^{(d-1)}} e(K) = e_1 + \dots + e_r,$$

where $\mathcal{S}_c^{(d-1)} = \{X_{d-1} \setminus L \mid L \in \mathcal{S}^{(d-1)}\}$.

Proof. Let us note that the product $d_1 \cdots d_r$ is equal to the r -th determinant divisor

$$\Delta_r(M) = \gcd\{\det M_{KS} \mid K \subset X_{d-1}, S \subset X_d, |K| = |S| = r\}.$$

Recall from Corollary 3.13 that $\det \partial_{KS} \neq 0$ if and only if $S \in \mathcal{S}^{(d)}$ and $X_{d-1} \setminus K \in \mathcal{S}^{(d-1)}$. Then, the exponent of $\Delta_r(M)$ is equal to the $\min_{K \in \mathcal{S}_c^{(d-1)}} e(K)$, and hence this leads to the formula. \square

Now, let us consider the $(d-1)$ -st persistent homology $H_{d-1}(\mathcal{X})$ and its lifetimes. Let p and q be the indices appearing in the indecomposable decomposition (2.5) for $H_{d-1}(\mathcal{X})$. Because of $\beta_{d-1}(X^{(d)}) = 0$, we have $q = 0$. Furthermore, it follows that $p \leq \dim \ker \delta_{d-1} = \text{rank } \delta_d = r$. In case of $p < r$, we add $l_i = 0$ for $i = p+1, \dots, r$ to the list of the lifetimes l_1, \dots, l_p .

Lemma 4.3. *$\{e_1, \dots, e_r\}$ and $\{l_1, \dots, l_r\}$ coincide as multisets.*

Proof. Let us express the boundary maps $C_d(\mathcal{X}) \xrightarrow{\delta_d} C_{d-1}(\mathcal{X}) \xrightarrow{\delta_{d-1}} C_{d-2}(\mathcal{X})$ in the matrix forms by using the standard bases Ξ_d, Ξ_{d-1} , and Ξ_{d-2} . Then, by performing appropriate base changes, δ_d is expressed as a smith normal form

$$\delta_d = \left[\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right], \quad A = \text{diag}(z^{e_1}, \dots, z^{e_r}),$$

where $e_i = t_{\sigma_i} - t_{\tau_i}$ is determined by the birth times of corresponding simplices $\sigma_i \in \Xi_d$ and $\tau_i \in \Xi_{d-1}$. We note that $t_{\sigma_i}, i = 1, \dots, r$, give the death times of the persistent homology $H_{d-1}(\mathcal{X})$. Furthermore, it follows from $\delta_{d-1} \circ \delta_d = 0$ that the first r columns of δ_{d-1} are now expressed to be zero vectors. It means that $t_{\tau_i}, i = 1, \dots, r$, are the birth times of $H_{d-1}(\mathcal{X})$, and hence $e_i, i = 1, \dots, r$, coincide with the lifetimes of $H_{d-1}(\mathcal{X})$. \square

Proof of Theorem 1.1. It follows from Lemma 4.2 and 4.3 that

$$L_{d-1} = l_1 + \dots + l_r = e_1 + \dots + e_r = \min_{K \in \mathcal{S}_c^{(d-1)}} e(K).$$

Note that the minimum of $e(K)$ is achieved by a minimum spanning acycle $L = X_{d-1} \setminus K \in \mathcal{S}^{(d-1)}$. Thus, by combining with Proposition 2.2, we obtain Theorem 1.1.

5. SIMPLICIAL COMPLEX PROCESS

5.1. Random Persistence Diagram as Point Process. First of all, we briefly recall the notion of point processes or random point fields. Let S be a locally compact Polish space (locally compact separable metrizable space) and $Q = Q(S)$ be the set of nonnegative integer valued Radon measures on S . Here, ξ is called a Radon measure if ξ is locally finite in the sense that $\xi(K) < \infty$ whenever $K \subset S$ is compact. Each element $\xi \in Q$ can be expressed as $\xi = \sum_s m_s \delta_s$, where δ_s is the delta measure at s and $m_s \in \mathbb{Z}_{\geq 0}$ stands for multiplicity.

Let $B_c(S)$ be the space of bounded measurable functions with compact support on S . For $f \in B_c(S)$, we define a coupling of $\xi \in Q$ and $f \in B_c(S)$ by

$$\langle \xi, f \rangle = \int_S f(s) \xi(ds) = \sum_s m_s f(s).$$

In particular, when f is the indicator function I_A of a measurable set A , $\langle \xi, I_A \rangle = \xi(A)$ is the number of points in A counting with multiplicity. A sequence $\{\xi_n \in Q\}_{n \geq 1}$ is said to converge to ξ *valuely* if $\langle \xi_n, f \rangle$ converges to $\langle \xi, f \rangle$ for any bounded continuous functions f with compact support. The space Q is equipped with the topological σ -algebra $\mathcal{B}(Q)$ with respect to the vague topology. A Q -valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a point process or a random point field.

Given a point process on S , the expectation $\lambda(A) := \mathbb{E}\xi(A)$ for every Borel set A defines a measure which may be finite or infinite. If it is also a Radon measure, λ is said to be the mean measure or the intensity measure. In this case, we have

$$\mathbb{E}[\langle \xi, f \rangle] = \int_S f(s) \lambda(ds)$$

for $f \in B_c(S)$. We note that the mean measure does not necessarily belong to Q . Higher moment measures can also be defined.

Let $\mathcal{X} = (X(t))_{t \in \mathbb{R}_{\geq 0}}$ be an increasing stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in the set of simplicial complexes, i.e., a random filtration of a simplicial complex. As in Section 2.2, we assume that there is a finite saturation time $T = T(\omega)$ such that $X(t) = X(T)$ for $t \geq T$ a.s.

As explained in Section 2, every filtration associates persistence diagrams on $\Delta = \{(x, y) \in \overline{\mathbb{R}}_{\geq 0}^2 \mid x \leq y\}$. Namely, a random filtration \mathcal{X} assigns a sequence of Q -valued random variables

$$\Xi = \{\xi_i \in Q(\Delta) \mid i \in \mathbb{Z}_{\geq 0}\},$$

where each ξ_i is the i -th persistence diagram of \mathcal{X} . In this case, the mean measure λ_i on Δ turns out to be a Radon measure (indeed a totally finite measure), and we call it the i -th mean persistence diagram. Hence, we have

$$\mathbb{E}[\langle \xi_i, f \rangle] = \int_{\Delta} f(x, y) \lambda_i(dx dy)$$

for $f \in B_c(\Delta)$, and it also makes sense for nonnegative measurable functions. In particular, this leads to

$$\mathbb{E}[L_i] = \int_{\Delta} (y - x) \lambda_i(dx dy)$$

for the lifetime sum L_i of the i -th persistent homology.

We consider two generalizations of the Erdős-Rényi graph process as random filtrations of simplicial complexes and discuss the expectation of the lifetime sum.

5.2. Linial-Meshulam Process. We discuss a stochastic process $\{\mathcal{K}^{(d)}(t)\}_{0 \leq t \leq 1}$ studied in [16]. Let Δ_{n-1} be the $(n-1)$ -dimensional maximal simplicial complex on the set $[n] = \{1, 2, \dots, n\}$, and let $\Delta_{n-1}^{(d)}$ be its d -dimensional skeleton ($1 \leq d \leq n-1$). Let $\{t_\sigma \mid \sigma \in (\Delta_{n-1})_d\}$ be i.i.d. random variables uniformly distributed on $[0, 1]$, where $(\Delta_{n-1})_d$ is the set of all d -simplices in Δ_{n-1} . We regard t_σ as the birth time of the d -simplex σ . Let $\{\mathcal{K}^{(d)}(t)\}_{0 \leq t \leq 1}$ be an increasing stochastic process on simplicial complexes defined by

$$\begin{aligned}\mathcal{K}^{(d)}(0) &= \Delta_{n-1}^{(d-1)}, \\ \mathcal{K}^{(d)}(t) &= \mathcal{K}^{(d)}(0) \sqcup \{\sigma \in (\Delta_{n-1})_d \mid t_\sigma \leq t\}.\end{aligned}$$

The process starts from the $(d-1)$ -dimensional skeleton $\Delta_{n-1}^{(d-1)}$ at time 0 and ends up with the d -dimensional skeleton $\Delta_{n-1}^{(d)}$ at time 1, i.e.,

$$\Delta_{n-1}^{(d-1)} = \mathcal{K}^{(d)}(0) \subset \mathcal{K}^{(d)}(t) \subset \mathcal{K}^{(d)}(1) = \Delta_{n-1}^{(d)}.$$

We call $\{\mathcal{K}^{(d)}(t)\}_{0 \leq t \leq 1}$ the d -Linial-Meshulam process. In particular, the 1-Linial-Meshulam process is nothing but the Erdős-Rényi graph process mentioned in Section 1.

Remark 5.1. Similar process is studied in [9], in which the birth times are i.i.d. exponential random variables with mean 1 instead of uniform random variables. The advantage of their choice of random birth times is that the process becomes a continuous-time Markov process.

Let $\beta_k(t)$ denote the k -th Betti number of $\mathcal{K}^{(d)}(t)$ at time t . Note that $\beta_k(t) = 0$ for $k = 0, 1, \dots, d-2$. We denote by $f_k(t) = f_k(\mathcal{K}^{(d)}(t))$ the number of k -simplices in $\mathcal{K}^{(d)}(t)$. Then, by applying the Euler-Poincaré formula to the d -Linial-Meshulam process, we have

$$(5.1) \quad \beta_d(t) - \beta_{d-1}(t) = f_d(t) - \binom{n-1}{d}.$$

We also note that there exist random times $\tau_{d-1}, T_d \in [0, 1]$ with $\tau_{d-1} \leq T_d$ such that

$$\begin{aligned}\beta_{d-1}(0) &= \binom{n-1}{d}, \quad \beta_{d-1}(t) = 0 \text{ for } t \geq \tau_{d-1}, \\ \beta_d(0) &= 0, \quad \beta_d(t) = \binom{n-1}{d+1} \text{ for } t \geq T_d.\end{aligned}$$

The Betti numbers $\beta_{d-1}(t)$ and $\beta_d(t)$ are non-increasing and non-decreasing in t , respectively.

5.3. Clique Complex Process. The clique complex $\text{Cl}(G)$ associated with a graph G is the maximal simplicial complex having G as the 1-dimensional skeleton. In other words, the simplices in $\text{Cl}(G)$ consist of all complete subgraphs in G . We define a clique complex process associated with the Erdős-Rényi graph process on n vertices by

$$\mathcal{C}(t) = \text{Cl}(\mathcal{K}^{(1)}(t)), \quad 0 \leq t \leq 1,$$

where $\mathcal{K}^{(1)}(t)$ is the one defined in the previous subsection. The process starts from the 0-skeleton, i.e., n isolated vertices, and ends up with Δ_{n-1} . Namely,

$$\Delta_{n-1}^{(0)} = \mathcal{C}(0) \subset \mathcal{C}(t) \subset \mathcal{C}(1) = \Delta_{n-1}.$$

By definition, for each edge e in $\mathcal{C}(t)$ (or equivalently $\mathcal{K}^{(1)}(t)$), a uniform random variable $t_e \in [0, 1]$ is independently assigned as its birth time, and the birth time of a simplex σ with $|\sigma| \geq 2$ is given by

$$t_\sigma = \max\{t_e \mid e \subset \sigma, |e| = 2\}.$$

We remark that $t_v = 0$ for each vertex $v \in [n]$.

Since a simplex σ contains $\binom{|\sigma|}{2}$ edges and t_σ is the maximum of the ordered statistics of i.i.d. $\binom{|\sigma|}{2}$ uniform random variables, we have

$$\mathbb{E}[t_\sigma] = \frac{\binom{|\sigma|}{2}}{\binom{|\sigma|}{2} + 1}.$$

Here, we used the following well-known fact. Let $y_i, i = 1, \dots, N$, be i.i.d. uniform random variables on $[0, 1]$ and $Y_i, i = 1, \dots, N$, be the rearrangement of y_i in increasing order. Then, for each $i = 1, \dots, N$,

$$(5.2) \quad \mathbb{E}[Y_i] = \frac{i}{N+1}.$$

We remark that the Betti numbers in $\{\mathcal{C}(t)\}_{0 \leq t \leq 1}$ are not monotone in t , although they are in $\{\mathcal{K}^{(d)}(t)\}_{0 \leq t \leq 1}$.

6. EXPECTATION OF LIFETIME SUM

In this section, we first prove Theorem 1.2. Then, we show a partial result on the expectation of the lifetime sum in the clique complex process. We note that, since both processes $\{\mathcal{K}^{(d)}(t)\}_{0 \leq t \leq 1}$ and $\{\mathcal{C}(t)\}_{0 \leq t \leq 1}$ are defined on the interval $[0, 1]$, the lifetime formula (1.5) is given as

$$(6.1) \quad L_{d-1} = \int_0^1 \beta_{d-1}(t) dt.$$

6.1. Proof of Theorem 1.2. For $d \geq 1$, let $\mathcal{C}_n^{(d)}$ be the set of d -dimensional simplicial complexes on n vertices with the $(d-1)$ -complete skeleton $\Delta_{n-1}^{(d-1)}$. For $Y \in \mathcal{C}_n^{(d)}$, let us define

$$\begin{aligned} \mathcal{R}_d(Y) &= \{\sigma \in (\Delta_{n-1})_d \mid \beta_{d-1}(Y \cup \sigma) = \beta_{d-1}(Y) - 1\}, \\ \mathcal{S}_d(Y) &= \{\sigma \in (\Delta_{n-1})_d \mid \beta_{d-1}(Y \cup \sigma) = \beta_{d-1}(Y)\}. \end{aligned}$$

We note that

- (1) $Y_d \subset \mathcal{S}_d(Y)$,
- (2) $(\Delta_{n-1})_d = \mathcal{R}_d(Y) \sqcup \mathcal{S}_d(Y)$ for $Y \in \mathcal{C}_n^{(d)}$, and
- (3) $\sigma \in \mathcal{S}_d(Y)$ is equivalent that the boundary of σ is contained in $\text{im } \partial_{Y,d}$,

where $\partial_{Y,d}$ is the d -th boundary map for Y .

The set $\mathcal{S}_d(Y) \setminus Y_d$ is called the shadow of Y in [17]. It should be noted that \mathcal{R}_d and \mathcal{S}_d are monotone decreasing and increasing, respectively, i.e.,

$$\mathcal{R}_d(Y) \supset \mathcal{R}_d(Y'), \quad \mathcal{S}_d(Y) \subset \mathcal{S}_d(Y')$$

for $Y, Y' \in \mathcal{C}_n^{(d)}$ with $Y \subset Y'$. For $Y \in \mathcal{C}_n^{(d)}$, we define the hull of Y by $\overline{Y} := Y \cup \mathcal{S}_d(Y)$. By definition, it is clear that

$$(6.2) \quad \beta_{d-1}(\overline{Y}) = \beta_{d-1}(Y).$$

Now we use a Kruskal-Katona-type result obtained in [17]. Here, we restate their result as to be fitted in our situation.

Proposition 6.1 ([17], Corollary 6.6). *Let Y be a d -dimensional simplicial complex with $|Y_d| = \binom{x}{d+1}$, where $x \geq d+1$ is a real. Then, $\text{rank } \partial_{Y,d} \geq \frac{d+1}{x}|Y_d|$. In particular, for any d -dimensional simplicial complex Y defined on n -vertices,*

$$(6.3) \quad \text{rank } \partial_{Y,d} \geq \frac{d+1}{n}|Y_d|.$$

Corollary 6.2. *For $Y \in \mathcal{C}_n^{(d)}$,*

$$(6.4) \quad \beta_{d-1}(Y) \leq \frac{d+1}{n}|\mathcal{R}_d(Y)|.$$

Proof. By (6.2) and (6.3),

$$\beta_{d-1}(Y) = \beta_{d-1}(\overline{Y}) = \binom{n-1}{d} - \text{rank } \partial_{\overline{Y},d} \leq \binom{n-1}{d} - \frac{d+1}{n}|\overline{Y}_d|.$$

Since $|\overline{Y}_d| = |\mathcal{S}_d(Y)| = \binom{n}{d+1} - |\mathcal{R}_d(Y)|$, we have the desired inequality. \square

In what follows, we will use the symbol N for $\binom{n}{d+1}$ in this subsection. Let us set $\mathcal{C}_{n,m}^{(d)} = \{Y \in \mathcal{C}_n^{(d)} \mid |Y_d| = m\}$. Then, we have a decomposition

$$\mathcal{C}_n^{(d)} = \bigcup_{m=1}^N \mathcal{C}_{n,m}^{(d)}.$$

Let $Y^{(d)}(n, m)$ be the uniform distribution on $\mathcal{C}_{n,m}^{(d)}$. We use the notation $Y \sim Y^{(d)}(n, m)$ to mean that Y is chosen according to the distribution $Y^{(d)}(n, m)$.

For two random simplicial complexes X and Y taking values in $\mathcal{C}_n^{(d)}$, we say that Y stochastically dominates X , denoted by $X \subset_{st} Y$, if there exists a coupling of X_d and Y_d such that $X_d \subset Y_d$ a.s.

Lemma 6.3. *Let $k, m \in \mathbb{N}$ with $km \leq N$. Suppose that $Y_1, \dots, Y_k \sim Y^{(d)}(n, m)$ are i.i.d. random simplicial complexes and $Y \sim Y^{(d)}(n, km)$. Then, $\cup_{i=1}^k Y_i \subset_{st} Y$.*

Proof. For given Y_1, \dots, Y_k , we define a collection of subsets of d -simplices by

$$\mathcal{A}_{Y_1, \dots, Y_k} := \{F \subset (\Delta_{n-1})_d \mid F \supset \cup_{i=1}^k (Y_i)_d, |F| = km\}.$$

We sample F from $\mathcal{A}_{Y_1, \dots, Y_k}$ uniformly at random and set $Y = \Delta_{n-1}^{(d-1)} \sqcup F$. Then, it is easy to see that the law of Y is equal to $Y^{(d)}(n, km)$, and hence $\cup_{i=1}^k Y_i \subset_{st} Y$. \square

For $Z \sim Y^{(d)}(n, m)$, we set $\rho_{n,m} = \mathbb{P}(\sigma \in \mathcal{R}_d(Z))$. By symmetry, the probability $\rho_{n,m}$ does not depend on the choice of $\sigma \in (\Delta_{n-1})_d$, and $\mathbb{E}|\mathcal{R}_d(Z)| = N\rho_{n,m}$. Note that $\rho_{n,m}$ is decreasing in m .

Lemma 6.4. *Let $k, m \in \mathbb{N}$ with $km \leq N$ and $Y \sim Y^{(d)}(n, km)$. Then,*

$$\mathbb{E}|\mathcal{R}_d(Y)| \leq N\rho_{n,m}^k.$$

Proof. Suppose $Y_1, \dots, Y_k \sim Y^{(d)}(n, m)$ are i.i.d. random simplicial complexes. From Lemma 6.3, we have a coupling such that $Y_i \subset \cup_{i=1}^k Y_i \subset Y$ for every $i = 1, 2, \dots, k$ by symmetry. Since \mathcal{R}_d is monotone decreasing, we obtain

$$\mathcal{R}_d(Y) \subset \mathcal{R}_d(\cup_{i=1}^k Y_i) \subset \cap_{i=1}^k \mathcal{R}_d(Y_i).$$

This implies

$$\begin{aligned} \mathbb{P}(\sigma \in \mathcal{R}_d(Y)) &\leq \mathbb{P}(\cap_{i=1}^k \{\sigma \in \mathcal{R}_d(Y_i)\}) \\ &= \mathbb{P}(\sigma \in \mathcal{R}_d(Y_1))^k \\ &= \rho_{n,m}^k. \end{aligned}$$

Therefore, again by symmetry, we obtain $\mathbb{E}|\mathcal{R}_d(Y)| \leq N\rho_{n,m}^k$. \square

Proposition 6.5. *Let $\{Y_t = \mathcal{K}^{(d)}(t)\}_{0 \leq t \leq 1}$ be the d -Linial-Meshulam process on n vertices. Then, for any $m \leq N$,*

$$\int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)| dt \leq \frac{m}{1 - \rho_{n,m}}.$$

Proof. For fixed $m \in \mathbb{N}$, we see that

$$\begin{aligned} \mathbb{E}|\mathcal{R}_d(Y_t)| &= \sum_{k=0}^{\lfloor N/m \rfloor} \sum_{\ell=km}^{(k+1)m-1} \mathbb{E}[|\mathcal{R}_d(Y_t)| \mid |(Y_t)_d| = \ell] \cdot \mathbb{P}(|(Y_t)_d| = \ell) \\ &= \sum_{k=0}^{\lfloor N/m \rfloor} \sum_{\ell=km}^{(k+1)m-1} \mathbb{E}|\mathcal{R}_d(Y^{(d)}(n, \ell))| \cdot \mathbb{P}(|(Y_t)_d| = \ell) \\ &\leq \sum_{k=0}^{\lfloor N/m \rfloor} \sum_{\ell=km}^{(k+1)m-1} \mathbb{E}|\mathcal{R}_d(Y^{(d)}(n, km))| \cdot \mathbb{P}(|(Y_t)_d| = \ell) \\ &\leq \sum_{k=0}^{\lfloor N/m \rfloor} N\rho_{n,m}^k \sum_{\ell=km}^{(k+1)m-1} \mathbb{P}(|(Y_t)_d| = \ell). \end{aligned}$$

Here $\mathcal{R}_d(Y^{(d)}(n, \ell))$ means $\mathcal{R}_d(Y)$ for $Y \sim Y^{(d)}(n, \ell)$. Since $|(Y_t)_d| \sim \text{Bin}(N, t)$, we have

$$\int_0^1 \mathbb{P}(|(Y_t)_d| = \ell) dt = \int_0^1 \binom{N}{\ell} t^\ell (1-t)^{N-\ell} dt = \frac{1}{N+1}.$$

Therefore,

$$\int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)| dt \leq \sum_{k=0}^{\lfloor N/m \rfloor} N\rho_{n,m}^k \frac{m}{N+1} \leq \frac{m}{1 - \rho_{n,m}}.$$

\square

Now, we appropriately choose m in Proposition 6.5. For $0 < c < 1$, let us define

$$(6.5) \quad m_c(n) := \min \left\{ m \leq N \mid \rho_{n,m} \leq c \right\}.$$

Then, Hoffman-Kahle-Paquette showed the following result.

Lemma 6.6 ([10], Lemma 10). $m_{1/2}(n) \leq 4\binom{n}{d}$.

We remark that a slightly different definition $m_{1/2}(n) = \min\{m \leq N \mid \mathbb{E}|\mathcal{R}_d(Y)| \leq N/2\}$ is used in [10], and it is equivalent to (6.5) with $c = 1/2$ by symmetry.

Proposition 6.7. *Let $\{Y_t\}_{0 \leq t \leq 1}$ be the d -Linial-Meshulam process on n vertices. Then,*

$$\int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)|dt \leq 8 \binom{n}{d}.$$

Proof. We set $m = m_{1/2}(n)$ in Proposition 6.5. Then, $\rho_{n,m} \leq 1/2$ from (6.5) and hence

$$\int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)|dt \leq \frac{m}{1 - \rho_{n,m}} \leq 2m_{1/2}(n) \leq 8 \binom{n}{d}.$$

□

Now we are in a position to prove our main result.

Proof of Theorem 1.2. By the lifetime formula (6.1), Corollary 6.2 and Proposition 6.7, we obtain an upper bound

$$\mathbb{E}[L_{d-1}] = \int_0^1 \mathbb{E}[\beta_{d-1}(t)]dt \leq \frac{d+1}{n} \int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)|dt \leq 8 \frac{d+1}{n} \binom{n}{d} \sim \frac{8(d+1)}{d!} n^{d-1}.$$

Let us next consider a lower bound. Because of $t_\eta = 0$ for any $(d-1)$ -simplex η in the d -Linial-Meshulam process, the lifetime formula (1.4) leads to

$$L_{d-1} = \text{wt}(T) \geq \sum_{i=1}^{|T|} u_i,$$

where T is the minimum spanning d -acycle and $0 \leq u_1 \leq u_2 \leq \dots \leq u_N \leq 1$ is the rearrangement of i.i.d. uniform random variables $\{t_\sigma \mid \sigma \in (\Delta_{n-1})_d\}$. We recall $|T| = \binom{n-1}{d}$ from Example 3.10. Then, it follows from (5.2) that

$$\mathbb{E}[L_{d-1}] \geq \sum_{i=1}^{|T|} \mathbb{E}[u_i] = \sum_{i=1}^{|T|} \frac{i}{N+1} \sim \frac{d+1}{2d!} n^{d-1}.$$

This completes the proof. □

6.2. Expected Lifetime Sum for Clique Complex Process. In this subsection, we consider the clique complex process and show bounds for the expectation of the lifetime sum. For the derivation of the upper bound, we first recall the discrete Morse theory.

Definition 6.8. Let X be a simplicial complex on a vertex set V . A partial matching consists of a partition of X into three sets A, Q , and K along with a bijection $\phi : Q \rightarrow K$ such that $\sigma \subset \phi(\sigma)$ and $|\phi(\sigma)| = |\sigma| + 1$ for each $\sigma \in Q$.

We denote a partial matching by $\mathcal{M} = (A, \phi : Q \rightarrow K)$. Given a partial matching, we set a relation \ll on Q by extending transitively the relation \triangleleft defined by

$$Q' \triangleleft Q \iff Q' \subset \phi(Q).$$

A partial matching \mathcal{M} is called an *acyclic matching* if \ll is a partial order. The elements in A are called *critical* simplices.

Theorem 6.9 ([6]). *Suppose X is a simplicial complex with an acyclic matching \mathcal{M} . Then, X is homotopy equivalent to a CW complex with exactly one k -cell for each critical k -simplex.*

Let us construct a $(d-1, d)$ -type acyclic matching $(A, \phi : Q \rightarrow K)$ as follows:

- (1) Suppose that the vertex set V is totally ordered as $1 < 2 < \dots < |V|$ and it induces the lexicographic order $<_{lex}$ on X .
- (2) For $\sigma \in X_{d-1}$, if there exists

$$\tau = \text{lexmin}\{\tilde{\tau} \in X_d \mid \sigma \subset \tilde{\tau}, \sigma <_{lex} \tilde{\tau}\},$$

then we add $\phi : \sigma \mapsto \tau$ as a pairing.

- (3) All the remaining simplices are set to be critical.

This is the acyclic matching used in [12].

Now we derive the following bounds of the expected lifetime sum in the clique complex process.

Theorem 6.10. *For the clique complex process $\{\mathcal{C}(t)\}_{0 \leq t \leq 1}$, there exist positive constants c and C (depending on d) such that, as $n \rightarrow \infty$,*

$$cn^{d-1} \leq \mathbb{E}[L_{d-1}] \leq Cn^{d-1} \log n$$

for $d = 1, 2$ and

$$cn^{\frac{(d+2)(d-1)}{2d}} \leq \mathbb{E}[L_{d-1}] \leq Cn^{d-1}$$

for $d \geq 3$.

We remark that $d - 1 = \frac{(d+2)(d-1)}{2d}$ for $d = 1, 2$.

Proof. Let us write $f_i(t) = f_i(\mathcal{C}(t))$ and $\beta_i(t) = \beta_i(\mathcal{C}(t))$. We recall the Morse inequality

$$\sum_{j=d-2}^d (-1)^{d-1-j} f_j(t) \leq \beta_{d-1}(t).$$

We observe that

$$\begin{aligned} \int_0^{\left(\frac{d+1}{n}\right)^{1/d}} \mathbb{E}[f_j(t)] dt &= \int_0^{\left(\frac{d+1}{n}\right)^{1/d}} \binom{n}{j+1} t^{\binom{j+1}{2}} dt \\ &\sim A_j^{(d)} n^{j+1 - \frac{1}{d} \{ \binom{j+1}{2} + 1 \}} \\ &= \begin{cases} A_j^{(d)} n^{\frac{(d+2)(d-1)}{2d}}, & j = d-1, d. \\ A_j^{(d)} n^{\frac{(d+2)(d-1)-2}{2d}}, & j = d-2. \end{cases} \end{aligned}$$

It is easy to check that $c_{d-1} := A_{d-1}^{(d)} - A_d^{(d)} > 0$ for every $d \geq 1$. Therefore, by (6.1) and the Morse inequality, we see that

$$\begin{aligned} \mathbb{E}[L_{d-1}] &\geq \int_0^{\left(\frac{d+1}{n}\right)^{1/d}} \mathbb{E}[\beta_{d-1}(t)] dt \\ &\gtrsim c_{d-1} n^{\frac{(d+2)(d-1)}{2d}}. \end{aligned}$$

This yields the lower bound since $\frac{(d+2)(d-1)}{2d} = d - 1$ for $d = 1, 2$.

For the upper bound, we use the discrete Morse theory. Let $f_{d-1}^*(t)$ be the number of critical $(d-1)$ -simplices in the $(d-1, d)$ -type acyclic matching for $\mathcal{C}(t)$. It follows from Theorem 6.9 that

$$\beta_{d-1}(t) \leq \min\{f_{d-1}(t), f_{d-1}^*(t)\}.$$

The expectation of the first term is given by $\mathbb{E}[f_{d-1}(t)] = \binom{n}{d} t^{\binom{d}{2}}$. On the other hand, we compute $\mathbb{E}[f_{d-1}^*(t)]$ as

$$\begin{aligned} \mathbb{E}[f_{d-1}^*(t)] &= \sum_{\sigma \in (\Delta_{n-1})_{d-1}} \mathbb{P}(\sigma \text{ is critical}) \\ &= \sum_{j=d}^n \sum_{1 \leq i_1 < i_2 < \dots < i_{d-1} < j} \mathbb{P}(\{i_1, i_2, \dots, i_{d-1}, j\} \text{ is critical}). \end{aligned}$$

A $(d-1)$ -simplex $\{i_1, i_2, \dots, i_{d-1}, j\}$ is critical if and only if the d -simplex $\{i_1, i_2, \dots, i_{d-1}, j, k\}$ does not appear for any $k \geq j+1$. Hence, we obtain

$$\begin{aligned} \mathbb{E}[f_{d-1}^*(t)] &= \sum_{j=d}^n \binom{j-1}{d-1} t^{\binom{d}{2}} (1-t^d)^{n-j} \\ &\leq \binom{n}{d-1} t^{\binom{d}{2}} \sum_{j=d}^n (1-t^d)^{n-j} \\ &= \binom{n}{d-1} t^{\binom{d}{2}-d}, \end{aligned}$$

and

$$\mathbb{E}[\beta_{d-1}(t)] \leq \min \left\{ \binom{n}{d} t^{\binom{d}{2}}, \binom{n}{d-1} t^{\binom{d}{2}-d} \right\}.$$

Therefore, we obtain

$$\mathbb{E}[L_{d-1}] \leq \int_0^1 \binom{n}{d-1} t^{\binom{d}{2}-d} dt = O(n^{d-1})$$

for $d \geq 3$ and

$$\begin{aligned} \mathbb{E}[L_{d-1}] &\leq \int_0^{n^{-1/d}} \binom{n}{d} t^{\binom{d}{2}} dt + \int_{n^{-1/d}}^1 \binom{n}{d-1} t^{\binom{d}{2}-d} dt \\ &= O(n^{\frac{(d-1)(d+2)}{2d}}) + O(n^{d-1} \log n) \\ &= O(n^{d-1} \log n) \end{aligned}$$

for $d = 1, 2$. This completes the proof. \square

7. CONCLUDING REMARKS

7.1. Limiting Constant. A detailed analysis of the Linial-Meshulam complex $\mathcal{K}^{(d)}(t)$ at time $t = c/n$ ($c \geq 0$) has recently been reported in [18]. By applying their results, we formally show that the limit $I_{d-1} := \lim_{n \rightarrow \infty} \frac{1}{n^{d-1}} \mathbb{E}[L_{d-1}]$ can be expressed by an integral form which recovers $I_0 = \zeta(3)$ for $d = 1$.

As was studied in the Erdős-Rényi graph, Poisson trees play an important role to characterize the Linial-Meshulam complex at $t = c/n$. By using the spectral measure of the upper $(d-1)$ -dimensional Laplacian $\partial_d \partial_d^T$ obtained from the boundary operator, Linial-Peled [18] basically show the following: let t_d^* be the unique root in $(0, 1)$ of the following equation

$$(d+1)(1-t) + (1+dt) \log t = 0$$

and set $c_d^* = \psi_d(t_d^*)$ by

$$\psi_d(t) = \frac{-\log t}{(1-t)^d}, \quad t \in (0, 1).$$

For $d = 1$, we understand $t_1^* = c_1^* = 1$. Let $t = t_c$ be the smallest positive root of the equation $t = e^{-c(1-t)^d}$. Then, for every $c > c_d^*$,

$$\frac{1}{\binom{n}{d}} \mathbb{E}[\beta_d(c/n)] = (1 + o(1)) \left\{ ct_c(1-t_c)^d + \frac{c}{d+1}(1-t_c)^{d+1} - (1-t_c) \right\}$$

holds with probability tending to 1 as $n \rightarrow \infty$.

Now, by applying this asymptotic formula into (5.1), we have

$$\begin{aligned} \frac{1}{\binom{n}{d}} \mathbb{E}[\beta_{d-1}(c/n)] &= \frac{1}{\binom{n}{d}} \left\{ \mathbb{E}[\beta_d(c/n)] + \binom{n-1}{d} - \mathbb{E}[f_d(c/n)] \right\} \\ &= (1 + o(1)) \underbrace{\left\{ ct_c(1-t_c)^d + \frac{c}{d+1}(1-t_c)^{d+1} + t_c - \frac{c}{d+1} \right\}}_{=: h_d(c)}. \end{aligned}$$

Then, integrating both sides roughly yields

$$\frac{1}{n^{d-1}} \int_0^1 \mathbb{E}[\beta_{d-1}(t)] dt \sim \frac{1}{d!} \binom{n}{d}^{-1} \int_0^n \mathbb{E}[\beta_{d-1}(c/n)] dc \approx \frac{1}{d!} \int_0^\infty h_d(c) dc$$

as $n \rightarrow \infty$. We should remark that the part \approx is the rough derivation. From this formal discussion, we conjecture that the limit exists and the limiting constant is given by

$$I_{d-1} = \lim_{n \rightarrow \infty} \frac{1}{n^{d-1}} \mathbb{E}[L_{d-1}] = \frac{1}{d!} \int_0^\infty h_d(c) dc.$$

Actually, this integral form recovers $I_0 = \zeta(3)$, which is nothing but Frieze's $\zeta(3)$ -limit theorem. Namely, for $d = 1$, we have $t_c = 1$ for $0 \leq c \leq 1$ and $t_c = \psi_1^{-1}(c)$ for $c \geq 1$. Then, we obtain

$$\begin{aligned} I_0 &= \int_0^1 \left(1 - \frac{c}{2}\right) dc + \int_1^\infty h_1(c) dc \\ &= \frac{3}{4} + \int_0^1 \frac{(2-2t+t \log t)(1-t+t \log t)}{2(1-t)^3} dt \\ &= \zeta(3). \end{aligned}$$

Here, the second equality follows by the change of variables $c = \psi_1(t)$ for $1 \leq c < \infty$.

7.2. Central Limit Theorem. After Frieze's work, Janson [11] proved the central limit theorem

$$\sqrt{n}(L_0 - \zeta(3)) \xrightarrow{d} N(0, \sigma^2)$$

with $\sigma^2 = 6\zeta(4) - 4\zeta(3)$. Hence, the next interesting problem is to find the variance of L_{d-1} , and furthermore to establish the (functional) central limit theorem in our setting. To this aim, we feel that we need more detailed study of structures of spanning acycles.

7.3. Limit Theorem of Persistence Diagram. In Section 5, we viewed random persistence diagrams as point processes on Δ . Then, the lifetime sum L_{d-1} is a functional of this point process and we derived its order in this paper. So, another natural problem is to establish the limit theorem for persistence diagrams. Moreover, limit theorems for barcodes and persistent landscapes for the Linial-Meshulam processes and the clique complex process should also be studied in connection with persistence diagrams.

7.4. Order in the Clique Complex Process. Theorem 6.10 shows the upper and lower bounds of $\mathbb{E}[L_{d-1}]$ in the clique complex process, but the explicit order has not yet been obtained at present. We performed numerical experiments to observe $\mathbb{E}[L_{d-1}]$ with respect to the number of vertices. From these computations, we observe that the upper bound seems to be correct for $d = 2$, although the lower bound is the right order for $d = 1$ (Frieze's $\zeta(3)$ -limit theorem).

7.5. Asymptotics of ℓ^2 -norm. In Remark 2.4, we showed the integral formula for the ℓ^2 -norm $\|\vec{l}\|_2$ of the sequence $\vec{l} = (l_i)_{i=1}^p$ of the lifetimes in connection with the persistence landscape. Then, it seems to be interesting to study asymptotic behaviors of $\|\vec{l}\|_2$ in a similar spirit to our main result. For this purpose, we would like to also derive an algebraic formulation of $\|\vec{l}\|_2$ corresponding to (1.4).

7.6. Wilson's Algorithm. The Wilson's algorithm [22] provides a fast algorithm of sampling uniform spanning trees by using loop-erased random walks on graphs. It would be natural to ask a generalization of the Wilson's algorithm producing uniform spanning acycles. To this aim, there are two things we need to consider. One is to find a natural candidate of loop-erased random walks defined on simplicial complexes. The other is to give a right meaning of *uniform* under the presence of the nontrivial factor (3.9) in (3.10). In case of graphs ($d = 1$), since this factor is always 1, the weight of each spanning tree is not biased.

ACKNOWLEDGEMENT

This work is partially supported by JSPS Grant-in-Aid (26610025, 26287019).

REFERENCES

- [1] P. Bubenik. Statistical topological data analysis using persistent diagram. arXiv:1207.6437v4.
- [2] C. Cooper, A. Frieze, N. Ince, S. Janson and J. Spencer. On the length of a random minimum spanning tree. arXiv:1208.5170v2.
- [3] A. Duval, C. J. Klivans and J. L. Martin. Simplicial matrix-tree theorems. *Trans. Amer. Math. Soc.* 361 (2009), 6073–6114.
- [4] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological Persistence and Simplification. *Discrete Comput. Geom.* 28 (2002), 511–533.
- [5] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen* 6 (1959), 290–297.
- [6] R. Forman. Morse theory for cell complexes. *Adv. Math.* 134 (1998), 90–145.
- [7] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Math.* 10 (1985), 47–56.
- [8] J. R. Munkres. Elements of Algebraic Topology. Perseus Publishing, 1984.
- [9] C. Hoffman, M. Kahle and E. Paquette. Spectral gaps of random graphs and applications to random topology. *Discrete Math.* 309 (2009), 1658–1671.
- [10] C. Hoffman, M. Kahle and E. Paquette. The threshold for integer homology in random d -complexes. arXiv:1308.6232v2.
- [11] S. Janson. The Minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Struct. Alg.* 7 (1995), 337–355.

- [12] M. Kahle. Topology of random clique complexes. *Discrete Math.* 309 (2009), 1658–1671.
- [13] M. Kahle. Topology of random simplicial complexes: a survey. arXiv:1301.7165v2.
- [14] G. Kalai. Enumeration of Q -acyclic simplicial complexes. *Israel J. Math.* 45 (1983), 337–351.
- [15] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.* 7 (1956), 48–50.
- [16] N. Linial and R. Meshulam. Homological connectivity of random 2-complexes. *Combinatorica* 26 (2006), 475–487.
- [17] N. Linial, I. Newman, Y. Peled and Y. Rabinovich. Extremal problems on shadows and hypercuts in simplicial complexes. arXiv:1408.0602v2.
- [18] N. Linial and Y. Peled. On the phase transition in random simplicial complexes. arXiv:1410.1281.
- [19] J. Little. A proof for the queuing formula: $L = \lambda W$. *Operations Research* 9 (1961), 383–387.
- [20] R. Lyons. Random complexes and ℓ^2 -Betti numbers. *J. Topology Anal.* 1 (2009), 153–175.
- [21] R. Meshulam and N. Wallach. Homological connectivity of random k -dimensional complexes. *Random Struct. Alg.* 34 (2009), 408–417.
- [22] D. B. Wilson. Generating random spanning trees more quickly than the cover time. Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996), 296303, ACM, New York, 1996.
- [23] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.* 33 (2005), 249–274.

INSTITUTE OF MATHEMATICS FOR INDUSTRY, KYUSHU UNIVERSITY, 744, MOTOOKA, NISHI-KU, FUKUOKA, 819-0395, JAPAN

E-mail address: `hiraoka@imi.kyushu-u.ac.jp`

INSTITUTE OF MATHEMATICS FOR INDUSTRY, KYUSHU UNIVERSITY, 744, MOTOOKA, NISHI-KU, FUKUOKA, 819-0395, JAPAN

E-mail address: `shirai@imi.kyushu-u.ac.jp`